Megiddo is Directed at Analysis of Association Rules for Statistical Significance and Not the Invention as Claimed

The Megiddo reference is directed to the process of analyzing generated association rules for statistical significance. (see col. 2, lines 62-64) Prior art methods for estimating statistical significance (e.g., chi-squared (col. 3, lines 11-12) and a support based method (col. 3, lines 14-18)) suffered from the drawback that they "did not take into account the number of hypotheses which were being tested." (col. 3, lines 13, 14, and 17) The Megiddo system apparently overcomes the drawback of these prior art approaches for estimating the statistical significance of association rules.

An example of the desirability of determining the significance of an association rule is set forth in col. 7, lines 30-67.

In this regard, the mining of association rules is only mentioned in passing. For example, set 102 of FIG. 3 states, "at step 102 in which association rules are mined from a database in a conventional manner.." (col. 9, line 62 to col. 10, line 8).

Step 104 of FIG. 3 evaluates the significance of each association rule in order to reduce (col. 7, lines 63-65) the total number of association rules (e.g., eliminate insignificant rules (col. 7, lines 2-3)). The disclosure

of Megiddo is primarily directed to describing step 104 and providing examples thereof.

RESPONSE TO ARGUMENTS SET FORTH ON PAGES 14-16 OF FINAL OFFICE ACTION

Pages 14 to 16 of the Action sets forth responses to arguments originally presented by Applicant in the 12/26/02 response to initial office action.

a) The Action recites col. 1 lines 59-60 to support the contention that Megiddo et al. uses continuous real-time information. First, the use of bar-code reader to read data from a product being purchased is common to most checkout systems. However, the mere fact that data is being collected in real-time does not imply and we cannot infer that the data is then processed in a time frame sufficient to capture dynamic trends.

A scalability challenge is described (page 2, lines 5-18) is in the Background of the patent application on page 4 lines 7 to 14.

> A second challenge is how to enable a conventional system, which is configured to process small amounts of data, to process very large data sets. In a conventional shopping network, a huge volume of transaction records must be processed everyday, and it is unlikely that centralized processing will yield satisfactory results. The scalability issue becomes more critical in the provision of real-time data mining service described above. In order to scale-up, a mechanism is needed to distribute data

processing, reduce data volumes at each local site by summarization, and mine data incrementally at multiple levels of aggregation. Unfortunately, the prior art does not provide a way to perform these tasks on very large data sets. [emphasis added]

As the Background of the patent application sets forth, the sheer volume of daily data (e.g., the data collected from all the sales registers of a national retail outlet) burdens most data mining applications so that the time it takes to process the current days transactions is longer than the time before new data (the next day's transaction data) arrives for processing. This ever-increasing backlog of unprocessed transactions is at the heart of the scalability problem.

The Megiddo reference does not appear to define or even identify this scalability problem. Furthermore, the Megiddo reference does not offer any solutions to this problem.

b. Col. 11, line 6 is relied on for teaching the processing of large amounts of data. It is respectfully noted that creating a large transaction database with millions of records is not the same as providing a system for continuously processing an ongoing flow of new data in such a manner to address the scalability problem outlined above.

c. Col. 3, lines 45-49 are relied on for teaching the step of generating the association cube, population cube, and base cube as claimed. Specifically, the Action equates "synthetic databases" with the recited cubes by arguing that "synthetic databases are essentially what cubes consists of at their core." Applicant respectfully disagrees.

Megiddo's Synthetic Databases Do Not Fairly Teach the Specific Data Cubes as Claimed

First, the synthetic database is not a data cube. The synthetic databases are generated from an original database (col. 7, lines 12-15). The original database is described as "one or more databases 20 and/or flat files (i.e., text files 22), which contain data about one or more consumer transactions." (col. 6, lines 36-38)

Second, it is noted that Meggido does not mention data cubes, which have different characteristics from databases and are accessed using different methods and tools. Furthermore, the server 14 is described as including a database system (e.g., a DB2 or ORACLE database)(col. 5, 51-54), which is not compatible with data cubes. Consequently, the database 20 and the synthetic databases created therefrom do not fairly teach or suggest the data cubes as claimed.

Moreover, even if the synthetic database could be construed as a data cube, which is not conceded, the synthetic database does not fairly teach or suggest the specific data cubes as claimed in claims 2-5 and examples of which are provided in figures 6 & 7 of the patent application.

Furthermore, the Megiddo reference fails to teach or suggest the specific steps or elements recited in the claims as described in greater detail hereinafter.

The process employed by Megiddo of generating synthetic databases does not teach or suggest the step of "deriving a confidence cube and a support cube of an association rule based on the association cube, population cube, and the base cube" as claimed.

With respect to the dependent claims, the Megiddo reference fails to teach or suggest the additional limitations recited therein.

For example, the Megiddo reference fails to teach or suggest the step of "generating a scoped association rule cube; wherein the step of deriving a confidence cube and a support cube of an association rule based on the association cube, population cube, and the base cube includes the step of deriving a confidence cube and a support cube of a scoped association rule based on the association cube, population cube, and the base cube," as claimed in claim 3.

Additionally, the Megiddo reference fails to teach or suggest the step of "generating an association rule with conjoint items cube," as claimed in claim 4.

Moreover, the Megiddo reference fails to teach or suggest the step of generating a functional association rule cube," as claimed claim 5.

It is unclear from the Final Rejection specifically how the Megiddo reference teaches or suggests these claimed features. It would be very helpful to the applicant and consistent with proper examination practice if the Action would indicate specific elements or portions within the applied references that teach specific claimed features.

Accordingly, a request is made herein for greater specificity so that the applicant may clearly understand what disclosures are being relied upon in the appropriate references to reject Claims 1 and claims depending thereon.

Furthermore, as noted in the previous response, the synthetic databases of Meggido are described in the following manner: "one or more smaller databases are generated which are subsets of the original database as described below using a random seed in order to resample the overall database." (see step 112 of FIG. 4 and col. 7, lines 12 to 15 of Meggido).

Col. 10 lines 18 to 52, further describe how the synthetic databases are generated. Specifically, col. 10 line 21 states "this process is also referred to as determining thresholds by resampling." The purpose of determining a threshold by resampling appears to be to "generate one or more synthetic data sets of transactions under a model where the occurrences of all of the items are independent (i.e., the transactions are generated independently)."

As is evident, the generation of synthetic databases (also referred to as "determining a threshold by resampling") is very different from the generation of an association cube, a population cube and a base cube based on the volume cube as claimed.

d. Col. 2, lines 41- 43 are cited to teach "a computational engine for mining and summarizing transaction data." Megiddo fails to teach or suggest OLAP-based computational engines for mining and summarizing transaction data.

e. Col. 4, lines 54 - 61 are cited to teach "a data warehouse that stores transaction data."

SYSTEM CLAIMS 8-16

With respect to independent claim 8, the Megiddo reference fails to teach or suggest the following claimed aspect:

a) a plurality of local stations having a local computation engine for mining and summarizing the local transaction data and for generating local customer profile cubes.

Furthermore, the Megiddo reference fails to teach or suggest b) at least one global station, coupled to the plurality of the local stations, the global station having a global computation engine for receiving the local customer profiles, merging and mining the local profile cubes, and generating global profile cubes and association rules based on said local profile cubes, and providing the global profile cubes and the association rules to said plurality of local stations, as claimed.

Moreover, the Megiddo reference fails to teach or suggest the specific elements recited in the claims depending upon independent claim 8.

Page 15, paragraph (d) of the Final Action asserts that col. 2, lines 41-43 teach all the elements as claimed in claim 8. However, the cited portion is merely a very general description of generating association rules for a set of transactions (D). It is noted that the cited portion, which appears in the Background of Megiddo, is a

general statement without any disclosure as to specific processing elements for mining association rules. Accordingly, this cited portion does not fairly teach or suggest the specific limitations recited in claim 8 directed to a plurality of local stations, a global station, and the specific operations performed by the recited elements.

For example, FIG. 1 of Megiddo only illustrates a well-known client-server architecture. In contrast, the invention as claimed has a very different structure and operates in a very different manner than that of Megiddo.

Megiddo neither suggests nor discloses a local customer profile cubes nor a global customer profile cube that is generated based on the plurality of local customer profile cube. Moreover, as described earlier, it is submitted that Megiddo does not fairly teach data cubes.

Referring to claim 9, Megiddo fails to teach or suggest one or more of the following claimed elements:

> wherein each of said plurality of LDOSs comprises a local data warehouse and at least one local OLAP server,
>
> the local data warehouse being adapted to receive and store said transaction data,
>
> wherein the local computation engine builds the local profile cubes that contains at least partial information regarding customer profiling by

periodically mining new transactions flowing into said local data warehouse and deriving patterns for local analysis, said local computation engine also being adapted to incrementally update said local profile cubes.

Referring to claim 10, Megiddo fails to teach or suggest one or more of the following claimed elements:

wherein said local data warehouse receives and stores transaction data in a first predetermined interval and wherein said local OLAP engine generates said local profile cubes in a second predetermined interval.

Referring to claim 11, Megiddo fails to teach or suggest one or more of the following claimed elements:

the global data warehouse for receiving and storing the local profile cubes, and

the global computation engine for combining summary information from each of said LDOSs to build and incrementally update said global profile cubes and association rules, and for providing feedback to said plurality of LDOSs.

It is unclear from the Final Rejection specifically how the Megiddo reference teaches or suggests these claimed elements. It would be very helpful to the applicant and consistent with proper examination practice if the Action would indicate specific elements or portions

within the applied Megiddo reference that teach specific claimed features. In this regard, if this rejection is maintained, greater specificity is respectfully requested so that the applicant may clearly understand what disclosures are being relied upon in the appropriate references to reject Claim 8 and claims depending thereon.

f. Col. 9, lines 56 - 59 are cited to teach "the step of mining and summarizing local and global transaction data." Specifically, "by summing this over all of the passes, we get Number cf Hypotheses <= (Number of Frequent Itemsets times Number of Frequent Items)," is cited for this teaching.

METHOD CLAIMS 17-25

Referring to Claim 17 and the claims dependent thereon, it is respectfully submitted that Megiddo fails to teach or suggest one or more of the following claim elements:

a) mining and summarizing, using a plurality of local servers ("LDOSs"), said transaction data to generate local profile cubes;

b) merging and mining, using at least one global server ("GDOS"), said local profile cubes received from said plurality of LDOSs to generate global profile cubes

and association rules based on said local profile cubes; and

c) feeding back said global profile cubes and association rules from said GDOS to said plurality of LDOSs for their business applications.

For example, col. 9, lines 56-59 (and also col. 13, line 52-55 and col. 9, lines 60-64 cited in an earlier Action) fail to teach or suggest the use of local servers and a global server to perform the specific steps as claimed.

Page 16 of the Final Office Action cites col. 9, lines 56-59 to support the limitations set forth in independent claim 17. First, the context of col. 9, lines 56-59 is the determination of "an upper bound on the number of hypotheses that are implicitly tested." (col. 9, lines 28-30) Megiddo appears to employ the upper bound to "estimate the number of false discoveries for a given threshold." (col. 9, lines 30-31) This discussion is set in the larger context of determining the "statistical significance of a set of associations." (col. 9, lines 5-6)

It is respectfully submitted that determining the "statistical significance of a set of associations," and "an upper bound on the number of hypotheses that are implicitly tested," does not fairly teach or suggest the specific processing steps recited in claim 17.

Consequently, it is unclear from the Final Rejection specifically how the Megiddo reference teaches or suggests these claimed processing steps. It would be very helpful to the applicant and consistent with proper examination practice if the Action would indicate specific elements or portions within the applied Megiddo reference that teach specific claimed features. In this regard, if this rejection is maintained, greater specificity is respectfully requested so that the applicant may clearly understand what disclosures are being relied upon in the appropriate references to reject Claim 17 and claims depending thereon.

Accordingly, it is respectfully submitted that the Megiddo, fails to teach or suggest the association rule generation method and system, as claimed.

In view of the foregoing, it is respectfully submitted that all pending claims of the present invention are now in condition for allowance. Reexamination and reconsideration of the pending claims are requested and allowance at an early date solicited. The Examiner is invited to telephone the undersigned if the Examiner has any suggestions, thoughts or comments, which might expedite the prosecution of this case.

Respectfully submitted,

Eric Ho

Eric Ho, Reg. No. 39,711
Attorney for Applicant
20601 Bergamo Way
Northridge, CA 91326

Dated: June 26, 2003
I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner of Patents and Trademarks, Washington, D.C. 20231 on the date below.

Eric Ho
Eric Ho (RN 39,711)

June 26, 2003
(Date)